

Comprehensive Basis-Set Testing of Extended Symmetry Adapted Perturbation Theory

Montgomery Gray and John M. Herbert*

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210 USA

(Dated: October 20, 2021)

Hybrid or “extended” symmetry-adapted perturbation theory (XSAPT) replaces traditional SAPT’s treatment of dispersion with better-performing alternatives, while at the same time extending two-body (dimer) SAPT to a many-body treatment of polarization, via a self-consistent charge-embedding procedure. The present work presents a systematic study of how total interaction energies and also energy components converge with respect to the choice of Gaussian basis set in XSAPT. Pople-style basis sets, while very efficient, consistently afford errors > 1 kcal/mol with respect to benchmark interaction energies for standard data sets of noncovalent complexes, whereas Dunning’s correlation-consistent basis sets and Karlsruhe basis sets perform much better. Hybrid treatments of dispersion afford much faster basis-set convergence as compared to traditional SAPT, and benchmark-quality results can be obtained using basis sets of triple- ζ quality, although the use of diffuse functions proves to be essential. The use of dimer Hartree-Fock calculations as a correction for higher-order induction can be performed in even smaller basis sets with negligible error, leading to a composite approach that offers speedups of $100\times$ even for small systems.

1 Introduction

Noncovalent interactions are ubiquitous in nature and hold many important functions in crystal packing,^{1–5} protein folding,^{6–10} host-guest binding in pharmaceuticals,^{11–14} as well as various materials science applications.^{15–17} Despite their prevalence, noncovalent forces are often misunderstood by chemists.^{18–20} The framework of symmetry-adapted perturbation theory (SAPT)^{20–26} offers an accurate and systematic *ab initio* approach to noncovalent interaction energies including an energy decomposition into physically-meaningful components: electrostatics, Pauli repulsion, induction, and dispersion.^{27–29} This decomposition is useful for obtaining physical insight that is backed by reliable *ab initio* calculations.^{11,19,20,30–36} The SAPT energy decomposition, which is inherently better separable than methods based on supramolecular density functional theory (DFT),^{37–39} can also be used to develop physically meaningful force fields.^{1,14,40,41}

The most widely used variant of SAPT is known as SAPT0,^{25,42} which combines Hartree-Fock (HF) wave functions for the isolated monomers (as zeroth-order states) with second-order perturbation theory for the intermolecular Coulomb potentials.^{22,25} When used in conjunction with Kohn-Sham (KS) density functionals with correct asymptotic behavior, intramolecular electron correlation can be incorporated by substituting the KS determinant for the HF one, in a method that we have called SAPT0(KS).^{20,42} (The use of asymptotically-correct functionals is crucial and the SAPT0 formalism should not be used with arbitrary density functionals.^{42–44}) Both traditional SAPT0 and SAPT0(KS) afford semiquantitative results at $\mathcal{O}(N^5)$

cost.^{25,42} The accuracy of both methods is limited by the accuracy of the dispersion interactions, which are not quantitative within the “uncoupled” second-order approximation^{45–47} that characterizes SAPT0, and which is similar to second-order Møller-Plesset perturbation theory (MP2).^{48,49}

To obtain quantitative accuracy, second-order dispersion must be replaced by better-performing alternatives. This has led to development of MP2 variants including MP2C⁴⁶ and MP2D,⁴⁷ and (within the SAPT formalism), DFT-SAPT.^{23,24} These methods are much more accurate than SAPT0 but retain that method’s $\mathcal{O}(N^5)$ cost. (For MP2C and DFT-SAPT, a density fitting approximation is required in order to obtain fifth-order scaling.^{50,51}) The “extended” (X)SAPT approach follows a similar strategy,²⁰ replacing second-order dispersion with either *ab initio* dispersion potentials (XSAPT + *aiD*)^{48,52–54} or else with a many-body dispersion model (XSAPT + MBD).^{55,56} The latter approach is currently the best-performing variant of the theory,^{20,55} and is the one used herein. Unlike the aforementioned methods, XSAPT incurs only $\mathcal{O}(N^3)$ cost, and it can be extended to clusters of molecules using a self-consistent charge embedding to capture non-additive polarization effects.^{54,56–59} For large systems, the monomer-based nature of XSAPT calculations makes this approach more affordable even than supramolecular DFT.^{52,54,55}

Some limited basis-set testing of XSAPT methods has been reported in previous work,^{42,54,55,60} but only for total interaction energies whereas as herein we examine convergence of individual energy components, which allows us to consider whether earlier tests may have benefited from error cancellation amongst energy components that may exhibit different convergence behavior. The basis-set dependence of traditional SAPT has also been carefully evaluated,²⁵ however the especially slow convergence of the dispersion energy in the traditional ap-

*herbert@chemistry.ohio-state.edu

proach means that those tests are not directly applicable to XSAPT. Moreover, because XSAPT was designed for large systems, we want to extend the basis-set testing beyond Dunning’s correlation consistent set and provided comprehensive tests for Karlsruhe and for Pople basis sets as well. These tests allow us to consider whether hybrid basis-set combinations (in which different energy components are evaluated in different basis sets) might further improve the efficacy of XSAPT calculations. Such hybrid approaches take advantage of the inherent separability of the SAPT or XSAPT interaction energy

2 Theory

The XSAPT formalism^{48,57,58} and the XSAPT + MBD method^{55,56} have been described in previous work, including a recent review.²⁰ These methods are briefly summarized below.

A. SAPT0(KS). The starting point for XSAPT is second-order SAPT0(KS) for a dimer, for which the decomposition of the total interaction energy (E_{int}) is expressed as^{48,54}

$$E_{\text{int}}^{\text{SAPT0}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} + \delta E_{\text{HF}} \quad (1)$$

Superscripts indicate orders in intermolecular perturbation theory but we drop these henceforth, setting $E_{\text{elst}} \equiv E_{\text{elst}}^{(1)}$ (electrostatics) and $E_{\text{exch}} \equiv E_{\text{exch}}^{(1)}$ (exchange or Pauli repulsion). Second-order contributions are summed to obtain the total induction energy

$$E_{\text{ind}} = E_{\text{ind}}^{(2)} + E_{\text{ind-disp}}^{(2)} \quad (2)$$

and the total dispersion energy

$$E_{\text{disp}} = E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} \quad (3)$$

The final term in eq. 1 is a so-called “ δE_{HF} ” correction for higher-order induction effects,^{20,22}

$$\delta E_{\text{HF}} = \Delta E_{\text{int}}^{\text{HF}} - (E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind,resp}}^{(2)} + E_{\text{exch-ind,resp}}^{(2)}) \quad (4)$$

Here, $\Delta E_{\text{int}}^{\text{HF}}$ is the counterpoise-corrected HF interaction energy and the second-order quantities $E_{\text{ind,resp}}^{(2)}$ and $E_{\text{exch-ind,resp}}^{(2)}$ are “response” analogues of the second-order terms in eq. 2, which require the solution of coupled-perturbed HF equations.⁶¹ For SAPT0(KS) calculations, the first- and second-order SAPT terms in eq. 4 should be computed at the HF level even if the corresponding terms in eq. 1 are computed based on KS orbitals.⁴² Notably, δE_{HF} is the only term in eq. 1 that requires iterations in

a dimer basis set, and therefore becomes a bottleneck for large monomers.

The accuracy of SAPT0(KS) interaction energies depends critically on the use of asymptotically-correct exchange-correlation functionals.^{20,42} Long-range corrected (LRC) functionals offer a simple means to enforce this constraint, but the range separation parameter ω must be tuned separately for each monomer in order to obtain correct asymptotics.^{42–44} To this end, we use the LRC- ω PBE functional⁶² combined with the “global density-dependent” (GDD) tuning procedure.^{42,63} Results in Ref. 42 show that GDD-tuned SAPT0(KS) and XSAPT + MBD results are essentially identical to those obtained using tuning based on the ionization energy criterion.⁴³

B. XSAPT. The XSAPT + MBD approach replaces the second-order dispersion energy in eq. 3 with a variant of the range-separated and self-consistently screened MBD model developed by Tkatchenko and co-workers.⁶⁴ (The original model of Ref. 64 must be modified at short range for use with SAPT, as described in Ref. 55.) In addition, self-consistent field (SCF) wave functions for the monomers are computed in the presence of wave function-derived embedding charges computed using the CM5 charge model,⁶⁵ as described in Ref. 56. (CM5 charges are based on the Hirshfeld atomic charge model procedure but introduce parameters in an effort to obtain better-quality dipole moments.) For a dimer system, this completes the specification of the XSAPT + MBD method.

For a system composed of more than two monomers, the XSAPT interaction energy is²⁰

$$E_{\text{int}}^{\text{XSAPT}} = \sum_{A,B>A} (E_{\text{elst}}^{AB} + E_{\text{exch}}^{AB} + E_{\text{disp}}^{AB} + E_{\text{ind}}^{AB} + \delta E_{\text{HF}}^{AB}) + E_{\text{pol}}^{\text{PW}} + E_{\text{pol}}^{\text{MB}} \quad (5)$$

where the summand in parentheses is the SAPT0(KS) interaction energy for dimer AB , meaning eq. 1 (with MBD replacing second-order dispersion) but without charge embedding. The terms outside of the pairwise sum in eq. 5 are the pairwise and many-body polarization energies. The former is defined as

$$E_{\text{pol}}^{\text{PW}} = \sum_{A,B>A} [E_{AB}^{\text{XSAPT}}(AB) - E_{AB}^{\text{SAPT}}] \quad (6)$$

where the term in square brackets is the difference between the charge-embedded energy for dimer AB and the SAPT energy computed without charge embedding. Finally, there is a many-body polarization energy

$$E_{\text{pol}}^{\text{MB}} = \sum_{A,B>A} [E_{AB}^{\text{XSAPT}}(AB \cdots N) - E_{AB}^{\text{XSAPT}}(AB)] \quad (7)$$

where $E_{AB}^{\text{XSAPT}}(AB \cdots N)$ is the energy of dimer AB embedded in an environment of atomic charges corresponding to the entire supersystem $AB \cdots N$ whereas $E_{AB}^{\text{XSAPT}}(AB)$ is the same dimer’s energy when embedding charges are included only on monomers A and B .

3 Computational Details

A. Density Functionals. For SAPT0(KS) calculations, and therefore for XSAPT, the range separation parameter in the LRC functional must be tuned individually for each monomer.^{42–44} We do this once and for all at the LRC- ω PBE/def2-TZVP level and then use these values in all subsequent calculations regardless of basis set. Previous work has shown that tuned values of ω are sensitive to the fraction of short-range exact exchange in the LRC functional but for a fixed functional (such as LRC- ω PBE in the present work), these values are rather insensitive to the choice of basis set.⁶⁶ Most calculations are reported here at the XSAPT + MBD level of theory⁵⁵ but some tests will be reported using XSAPT + *aiD3*,⁵⁴ where “*aiD3*” denotes the third-generation *ab initio* dispersion potential.

B. Basis Sets. A standard complement of Pople-style basis sets is tested, ranging from 6-31G to 6-311++G(3df 2dp).⁶⁷ These are often regarded as low-quality basis sets in modern electronic structure theory but are still widely used, in part because some codes have been optimized to take advantage of their “*sp*” basis functions, meaning *s* and *p* functions with a common orbital exponent. All calculations in this work were performed using Q-Chem v. 5,⁶⁸ for which the use of Pople basis sets is significantly more efficient as compared to other basis sets with a comparable number of functions. Karlsruhe (or Ahlrichs) “def2” basis sets up to quadruple- ζ quality are also tested,^{69,70} as are Dunning’s correlation-consistent basis sets cc-pVXZ⁷¹ and aug-cc-pVXZ,⁷² for $X = \text{D, T, and Q}$.⁶⁷

In addition to these three standard families of basis sets, we also tested “calendar” versions of the correlation consistent basis sets, in which diffuse functions are systematically removed starting from aug-cc-pVXZ.⁷³ The jul-cc-pVXZ basis sets consist of cc-pVXZ for hydrogen and aug-cc-pVXZ for other atoms (meaning that all diffuse functions are removed from hydrogen), and jun-cc-pVXZ additionally removes the set of diffuse functions with highest angular moment from each non-hydrogen atom. (These basis sets were added to Q-Chem as part of the present work.) We note that jun-cc-pVDZ is suggested as a compromise basis set in SAPT0 calculations,^{25,42} exploiting the slow basis-set convergence of the dispersion energy to limit the intrinsic overestimation of dispersion (at the level of second-order perturbation theory) through the use of a limited ba-

sis set. This is a compromise because electrostatics is not generally converged at the double- ζ level,^{42,54} but overall errors in SAPT0 interaction energies are worse in aug-cc-pVDZ and larger basis sets, as compared to jun-cc-pVDZ.²⁵

We also modified the Karlsruhe basis sets to delete diffuse functions in an analogous manner, although these basis sets contain fewer diffuse functions as compared to the Dunning-style basis sets, *e.g.*, def2-SVPD for second-row atoms contains a diffuse *s* function and a set of diffuse *d* functions, but no diffuse *p* function, and for hydrogen there is a set of diffuse *p* functions but no diffuse *s* function.⁷⁰ As a first truncation, and in analogy to to jul-cc-pVXZ, we delete the diffuse functions on hydrogen. This results in what has historically been termed a “heavy-augmented” basis set, and we denote these, *e.g.*, as “ha-def2-SVPD” for the basis set that starts from def2-SVPD. As a second step, and in analogy to jun-cc-pVXZ, we delete the highest angular momentum set of diffuse functions on each non-hydrogen atom. For the second row, this leaves only minimal augmentation with a diffuse *s* function and thus we refer to these basis sets as “ma-def2”. This is similar in spirit to the partially-augmented Karlsruhe basis sets that were constructed in Ref. 74, although the diffuse exponents differ because Ref. 74 did not start from the standardized Karlsruhe diffuse exponents that were introduced in Ref. 70, and which do serve as the starting point for our truncated ha-def2 and ma-def2 basis sets.

C. Data Sets. For high-throughput evaluation of a wide variety of basis sets, we will use the S66 data set of noncovalent dimers.⁷⁵ This database was developed to sample various types of noncovalent interactions and benchmark interaction energies were reported in Ref. 75 at the level of coupled-cluster theory with single, double, and perturbative triple excitations [CCSD(T)], extrapolated to the complete basis set (CBS) limit. Per the analysis in Ref. 75, The 66 dimers in this test set are divided into three subsets: hydrogen-bonded dimers, which are characterized by the condition $|E_{\text{elst}}| \geq 2|E_{\text{disp}}|$; dispersion-dominated dimers for which $|E_{\text{disp}}| \geq 2|E_{\text{elst}}|$; and dimers of mixed-influence interactions where neither of these conditions are met. The hydrogen-bonded subset, which consists of 23 different dimers involving water, methanol, acetic acid, and other polar monomers places stringent demands on the basis set as compared to the other S66 systems, so it is useful to group the complexes in this way. The dispersion-bound subset contains 23 dimers involving monomers such as benzene, pyridine, and ethene.

We will also consider three data sets containing ionic monomers:⁷⁶ AHB21, which consists of 21 anion–neutral hydrogen-bonded complexes with ions including F^- , Cl^- , N_3^- , and SH^- ; CHB6, consisting of three alkali–benzene and three alkali–water cation–neutral complexes; and finally IL16, which is a set of 16 ion pairs representing

Table 1: Error Statistics for the SAPT0 Method Applied to the S66 Data Set.

Basis Set	Error (kcal/mol)	
	MAE	Max
jun-cc-pVDZ	0.51	1.55
jul-cc-pVDZ	0.64	2.34
aug-cc-pVDZ	0.67	2.49
jun-cc-pVTZ	0.81	3.34
jul-cc-pVTZ	0.93	3.56
aug-cc-pVTZ	1.01	3.70
jun-cc-pVQZ	1.05	3.91
jul-cc-pVQZ	1.09	3.98

constituent molecules or constituent moieties of ionic liquids. Benchmark interaction energies at the CCSD(T)/CBS level are taken from Ref. 76 for all three of these data sets.

4 Results and Discussion

A. Tests of Traditional SAPT0 Using S66. The convergence of traditional SAPT methods (including SAPT0) has been reported previously²⁵ but we include our own SAPT0 convergence tests here because they establish a baseline to which we can later compare the XSAPT methods, whose alternative descriptions of dispersion lead to accelerated basis-set convergence. Table 1 provides the mean absolute errors (MAEs) for the S66 database at SAPT0/cal-cc-pVXZ levels of theory, where X = D, T, or Q and cal = jun, jul, or aug. As previously reported,²⁵ the jun-cc-pVDZ basis set gives the best results with a MAE of 0.5 kcal/mol. Errors increase both with increasing cardinality of the basis set (*i.e.*, double-, triple-, or quadruple- ζ) and with increasing augmentation. The uncoupled or MP2-like second-order approximation for dispersion that is used in SAPT0 tends to overestimate dispersion significantly, yet converges very slowly to the CBS limit. More complete basis sets therefore afford increasingly poor dispersion energies and the use of jun-cc-pVDZ is a compromise that balances slow convergence against overestimation of the result. It is a remarkably robust compromise in small molecules (as in the S66 data set),^{25,42} although it may fare worse in larger systems. For example, in the L7 set of large dispersion-bound complexes,⁷⁷ the MAE for SAPT0/jun-cc-pVDZ interaction energies is 4.8 kcal/mol (versus 0.5 kcal/mol for S66), and the maximum error is 10.3 kcal/mol.⁴⁸

SAPT0 dispersion energies for all of the S66 dimers are plotted in Fig. 1 in a variety of basis sets. These data clearly demonstrate that there are nontrivial changes between triple- and quadruple- ζ basis sets. With the exception of dispersion, however, the other energy components

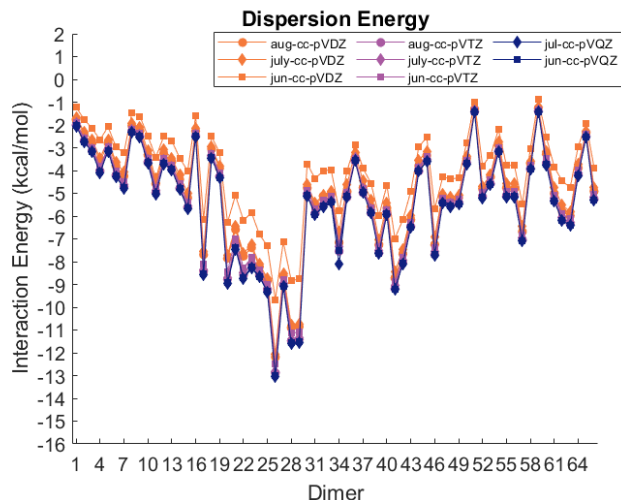


Fig. 1: Dispersion energies for the S66 dimers computed using SAPT0 with a variety of correlation-consistent basis sets and “calendar” variants thereof.

are essentially constant across these basis sets.

B. Broad Survey of Basis Sets for XSAPT + MBD Using S66. We next consider the performance of the hybrid XSAPT + MBD method as applied to the S66 dimers. These systems are small (with the largest being pentane dimer) and thereby facilitate high-throughput testing. XSAPT + MBD errors for each of the S66 dimers, across a wide range of Pople, Karlsruhe, and Dunning basis sets are plotted in Fig. 2 and will be analyzed below. A statistical survey of the results is presented in Table 3, broken down into the three subsets that were described in Section 3C. As with the SAPT0 assessment in Section 4A, error is measured relative to CCSD(T)/CBS benchmarks.⁷⁵

1. Karlsruhe Basis Sets

Errors for XSAPT + MBD using Karlsruhe “def2” basis sets are plotted in Fig. 2a, where the double-, triple-, and quadruple- ζ basis sets have been grouped together by color. Clearly, the double- ζ errors are much larger, exceeding 1 kcal/mol in many cases and typically 1–3 kcal/mol larger than what is obtained in more complete basis sets.

Figure 2a also demonstrates the importance of including diffuse functions when calculating interaction energies. This can be seen most clearly from the double- ζ data but is true as well in triple- and quadruple- ζ basis sets, although the effect is smaller in those larger basis sets. Errors increase in a consistent way as the diffuse orbitals are trimmed, going from def2-SVPD (with a full complement of diffuse functions) to the “heavy-augmented” ha-def2-SVPD basis set, and finally to the

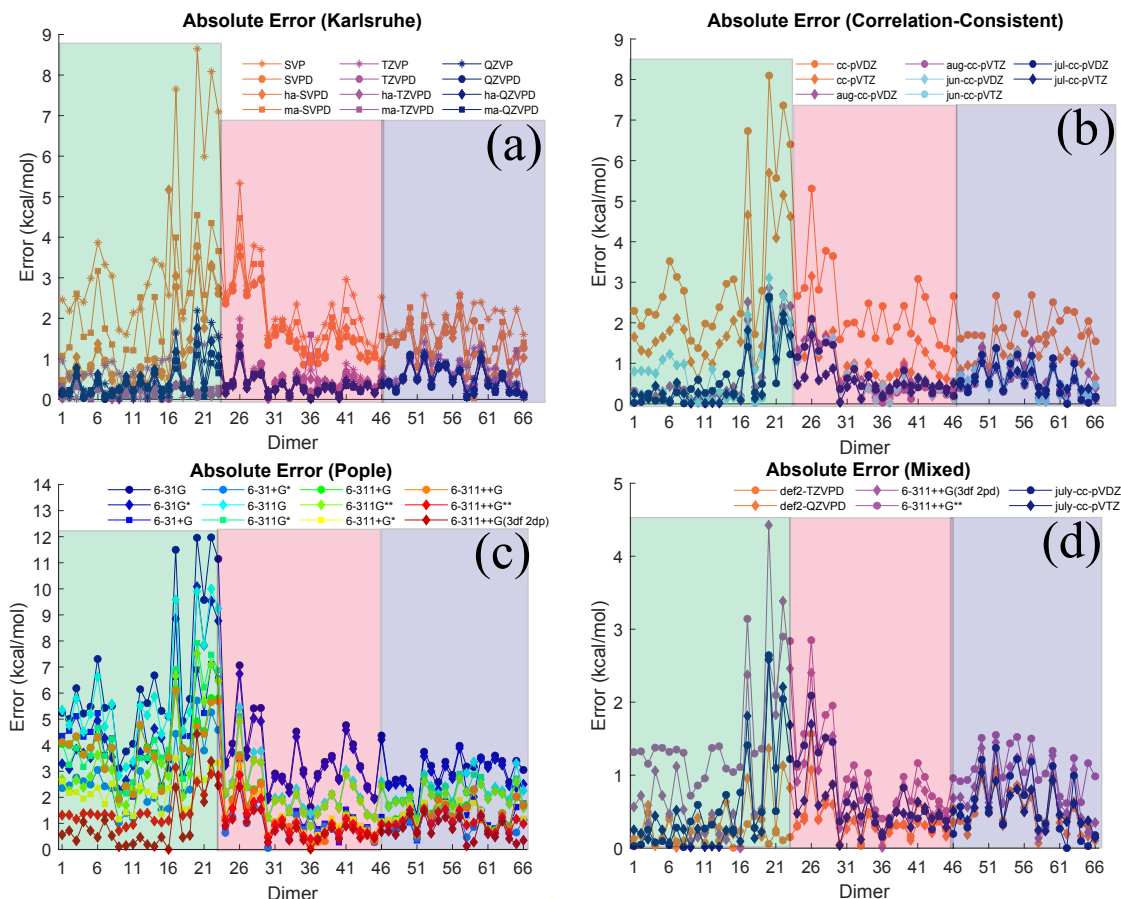


Fig. 2: Absolute errors in XSAPT+MBD total interaction energies for the S66 dimers, using (a) Karlsruhe basis sets, (b) Dunning correlation-consistent basis sets, (c) Pople basis sets, and (d) the best-performing basis sets from amongst these three categories. Indices along the horizontal axis refer to the way that the dimers were ordered in the original work of Ref. 75 and the three color-coded regions delineate the hydrogen-bonded subset (dimers 1–23), the dispersion-bound subset (24–46), and the subset of mixed-influence dimers (47–66). Errors are defined with respect to the CCSD(T)/CBS benchmarks in Ref. 75.

“minimally-augmented” ma-def2-SVPD basis. The importance of diffuse basis functions is most significant in systems that are dominated by hydrogen bonding, where induction effects are important, and removal of the diffuse functions has a smaller effect in systems that are dominated by dispersion.

Most interesting is that the absolute errors converge at the triple- ζ level. Difference between triple- and quadruple- ζ interaction energies are uniformly smaller than 1 kcal/mol and on average these differences are less than 0.1 kcal/mol. The largest differences (approaching 1 kcal/mol) are for systems with very strong hydrogen bonds, such as acetic acid dimer, and in these cases the quadruple- ζ errors are actually larger than the triple- ζ errors. Another important feature is that removal of the diffuse functions has a much smaller effect at the triple- and quadruple- ζ level than it does at the double- ζ level where these functions are absolutely essential for obtaining quantitative or semi-quantitative results. In the double- ζ basis sets, it seems that the diffuse functions are partially compensating for the overall incompleteness of

the basis set, which becomes unnecessary at the triple- ζ level.

2. Dunning Basis Sets

Errors for correlation-consistent basis sets and “calendar” variants thereof are provided in Fig. 2b. In the absence of any diffuse functions, even the cc-pVTZ basis set affords unacceptably large errors, specifically for the hydrogen-bonded complexes. These are substantially reduced, even at the double- ζ level, by minimal augmentation, *e.g.*, jun-cc-pVDZ or jun-cc-pVTZ, although some of the errors for hydrogen-bonded complexes remain larger than 1 kcal/mol even at the aug-cc-pVTZ level, where the average error is < 0.5 kcal/mol. The takeaway from this analysis is that the accuracy obtained from these very large basis sets is comparable to what is possible using the somewhat more modest Karlsruhe basis sets.

Table 2: Error Statistics for XSAPT + MBD Applied to the S66 Data Set and Subsets Thereof.

Method	Error (kcal/mol)							
	H-Bonded		Dispersion		Mixed		Total	
	MAE	Max	MAE	Max	MAE	Max	MAE	Max
def2-SVP	3.71	8.65	2.32	5.33	1.89	2.60	2.67	8.65
def2-SVPD	1.17	3.78	1.65	3.54	1.33	2.30	1.39	3.78
ha-def2-SVPD	1.35	5.17	1.75	3.74	1.22	2.06	1.45	5.17
ma-def2-SVPD	2.13	4.54	1.98	4.47	1.45	2.56	1.87	4.54
def2-TZVP	0.75	1.67	0.55	1.99	0.82	1.42	0.70	1.99
def2-TZVPD	0.17	0.48	0.43	1.57	0.54	0.99	0.37	1.57
ha-def2-TZVPD	0.21	0.57	0.50	1.56	0.55	0.99	0.41	1.57
ma-def2-TZVPD	0.19	0.36	0.51	1.78	0.60	1.17	0.43	1.78
def2-QZVP	0.53	2.19	0.36	1.32	0.55	1.17	0.48	2.19
def2-QZVPD	0.40	1.36	0.33	1.07	0.49	1.10	0.40	1.36
ha-def2-QZVPD	0.41	1.10	0.34	1.09	0.49	1.07	0.41	1.10
ma-def2-QZVPD	0.50	1.75	0.33	1.33	0.48	1.06	0.44	1.75
cc-pVDZ	3.34	8.10	2.42	5.31	1.90	2.68	2.58	8.10
cc-pVTZ	2.18	5.69	1.08	3.15	1.21	1.83	1.50	5.69
aug-cc-pVDZ	0.75	2.86	0.77	2.06	0.74	1.52	0.75	2.86
aug-cc-pVTZ	0.60	2.65	0.42	1.60	0.48	1.04	0.50	2.65
jul-cc-pVDZ	0.58	2.64	0.73	2.09	0.62	1.37	0.64	2.64
jul-cc-pVTZ	0.56	2.58	0.52	1.70	0.54	1.03	0.54	2.58
jun-cc-pVDZ	0.98	3.10	0.69	2.09	0.65	1.32	0.78	3.10
jun-cc-pVTZ	0.54	2.51	0.54	1.83	0.53	1.02	0.53	2.51
6-31G	6.61	11.97	3.56	7.06	3.02	3.98	4.46	11.97
6-31G(d)	4.74	10.08	3.41	6.74	2.73	3.88	3.67	10.08
6-31+G	4.46	7.11	1.20	3.41	1.25	1.98	2.35	7.11
6-31+G(d)	2.71	5.71	1.00	2.88	0.99	1.65	1.59	5.71
6-311G	5.84	10.00	2.26	5.44	2.28	3.36	3.52	10.00
6-311G(d)	4.23	7.92	2.13	5.12	2.19	2.90	2.88	7.92
6-311+G	3.95	6.16	1.16	3.62	1.42	2.30	2.21	6.16
6-311G(d,p)	3.54	7.51	2.11	4.91	2.02	2.71	2.58	7.51
6-311+G(d)	2.20	3.51	1.15	3.25	1.31	1.91	1.56	3.51
6-311++G	3.91	6.09	1.08	3.55	1.32	2.29	2.14	6.09
6-311++G(d,p)	1.53	3.14	1.02	2.85	1.09	1.55	1.22	3.14
6-311++G(3df 2dp)	0.99	4.43	0.79	2.40	0.77	1.15	0.85	4.43

3. Pople Basis Sets

XSAPT + MBD absolute errors using Pople basis sets are shown in Fig. 2c and are color-coded (across the visible spectrum) based on the size of the basis set, with warmer colors (towards red) representing larger basis sets. Errors are generally larger than when the Karlsruhe or Dunning basis sets are used, although the 6-311++G(3df,2dp) basis set (which was originally developed for MP2 calculations⁷⁸) performs reasonably well and best amongst the Pople-style basis sets that are tested here. That said, outside of the hydrogen-bonded subset of S66, where large induction effects necessitate the use of a considerable number of polarization functions, the performance of 6-311++G(3df,2dp) is nearly identical to that of 6-311++G(d,p), 6-311+G(d), 6-

311++G, 6-31+G(d), and even 6-31+G. A set of diffuse functions is necessary even for the dispersion-bound complexes but otherwise the double- ζ basis sets perform just as well as the triple- ζ ones when hydrogen bonds are absent. Polarization functions matter little (< 0.1 kcal/mol) for dispersion-bound systems.

Overall, even the better-quality Pople basis sets afford larger errors than what is possible to achieve readily with either Karlsruhe or Dunning basis sets, but they can be competitive when sufficient diffuse and polarization functions are added. In particular, these tests once again highlight the importance of diffuse functions especially in the context of hydrogen bonding. The use of composite *sp* shells makes Pople basis sets significantly faster (per unit basis function) as compared to other types of basis sets, assuming that one is using a quan-

tum chemistry program that is written to exploit this. This means that basis sets such as 6-311++G(d,p) and 6-311++G(3df,2dp) may have a place in the pantheon of SAPT methods for large systems.

4. Comparison Between Subsets of Basis Sets

Several of the best-performing basis sets from each of the categories discussed above are compared in Fig. 2d. It is clear that 6-311++G(d,p) is outperformed by both Dunning and Karlsruhe basis sets, with absolute errors that are larger by up to 2 kcal/mol. All of these worst-case discrepancies are found amongst the hydrogen-bonded subset of S66, hence the problem is likely the inadequate description of dispersion energies. (The basis-set behavior of individual energy components is considered in Section 4 C.)

In contrast, the best of the Dunning and Karlsruhe basis sets are nearly identical in their performance, with MAEs < 0.5 kcal/mol. The Karlsruhe basis sets achieve this level of accuracy with fewer basis functions and for that reason we will focus on the Karlsruhe basis sets in much of the rest of this paper.

C. Energy Component Analysis for S66. Figure 3 shows the comparison of the four energy components (electrostatics, exchange, induction, and dispersion) across the S66 data set, computed at the XSAPT + MBD level in various basis sets, although limited to the ones that perform reasonably well for total interaction energies, based on the analysis presented above. These data suggest that the basis sets that we singled out as the best-performing ones achieve this status not through any kind of error cancellation but rather because they offer a quantitative (or nearly quantitative) description of each of the energy components.

Dispersion energies, computed according to the MBD model,⁵⁵ are plotted in Fig. 3a and are essentially identical across all basis sets tested, including triple- ζ Pople basis sets such as 6-311++G(d,p). This behavior is not entirely surprising given that the dispersion model is based upon the SCF monomer electron densities, which are mapped onto a harmonic oscillator Hamiltonian by means of a Hirshfeld partition of the density into atomic contributions.⁶⁴ Convergence of SCF charge densities is usually achieved at the triple- ζ level and the Hirshfeld weighting scheme is also density-based. The behavior of MBD stands in marked contrast to the slow convergence of perturbative dispersion within the SAPT formalism, which is slower than that of other energy components; this was a major part of the original motivation for the development of hybrid XSAPT methods that treat dispersion differently.⁵²

For the electrostatic energy (Fig. 3b), the 6-311++G(d,p) basis set is not quite sufficient to obtain converged results for strong hydrogen bonds, where

it tends to exaggerate the electrostatic energy. (This was seen previously in tests for anion–water clusters.⁶⁰) However, other good-performing basis sets, including 6-311++(3df,2pd), appear to be converged in the sense that the electrostatic energies are indistinguishable from results obtained in the def2-QZVPD basis set.

Exchange energies (Fig. 3c) also exhibit fairly small discrepancies amongst the best-performing basis sets, with 6-311++G(d,p) tending to underestimate the repulsiveness of E_{exch} , by up to about 4 kcal/mol. This leads to some error cancellation in total interaction energies, given that the same basis set also exaggerates the attractiveness of E_{elst} . (Again, this is consistent with previous experience for hydrated anions.⁶⁰) The exchange energy, equivalent to Pauli repulsion between filled orbitals, generally increases in magnitude with the completeness of the basis, as the tails of the density become better described. The Karlsruhe and Dunning basis sets generally afford comparable values of E_{exch} except that the former exhibits slightly larger exchange energies for the strongest hydrogen bonds.

Finally, the induction energies (Fig. 3d) are only sensitive to basis set in the case of strong hydrogen bonds. In those cases, def2-QZVPD and jul-cc-pVTZ afford similar results and 6-311++G(3df,2pd) is also similar, but 6-311++G(d,p) lacks sufficient polarization functions to describe the polarization of the monomer densities and thus underestimates E_{ind} .

Consistent with the analysis of Section 4 B, we conclude that 6-311++G(d,p) is not fully converged but that 6-311++G(3df,2dp) offers performance that is comparable to the best Karlsruhe and Dunning basis sets. Moreover, it does so in a manner that does not rely on error cancellation.

D. Performance for Data Sets Containing Ions.

Ionic dimers have much larger interaction energies than neutral systems and may place different demands on basis sets. We next examine the performance for the IL16, AHB21, and CHB6 data sets,⁷⁶ with error statistics reported in Table 4. Because our testing for S66 revealed that MBD values of E_{disp} converge rapidly with the density, these tests are carried out using the XSAPT + *ai*D3 method,⁵⁴ where the dispersion energy (computed via atom–atom C_6 and C_8 potentials) does not depend on the density. In order to eliminate spurious charge transfer the monomer basis was used in the XSAPT + *ai*D3 calculations, although the dimer basis is still used for the δHF calculations as $\Delta E_{\text{int}}^{\text{HF}}$ in eq. 4 must be counterpoise-corrected. Table 3 lists the error statistics, with respect to CCSD(T)/CBS benchmarks, for XSAPT + *ai*D3 in various basis sets. Because the IL16 and AHB21 data sets contain anions, only basis sets that are at least minimally augmented with diffuse functions are considered for these benchmarks.

The first column of Table 3 gives the error statistics for the IL16 database of cation–anion pairs. The

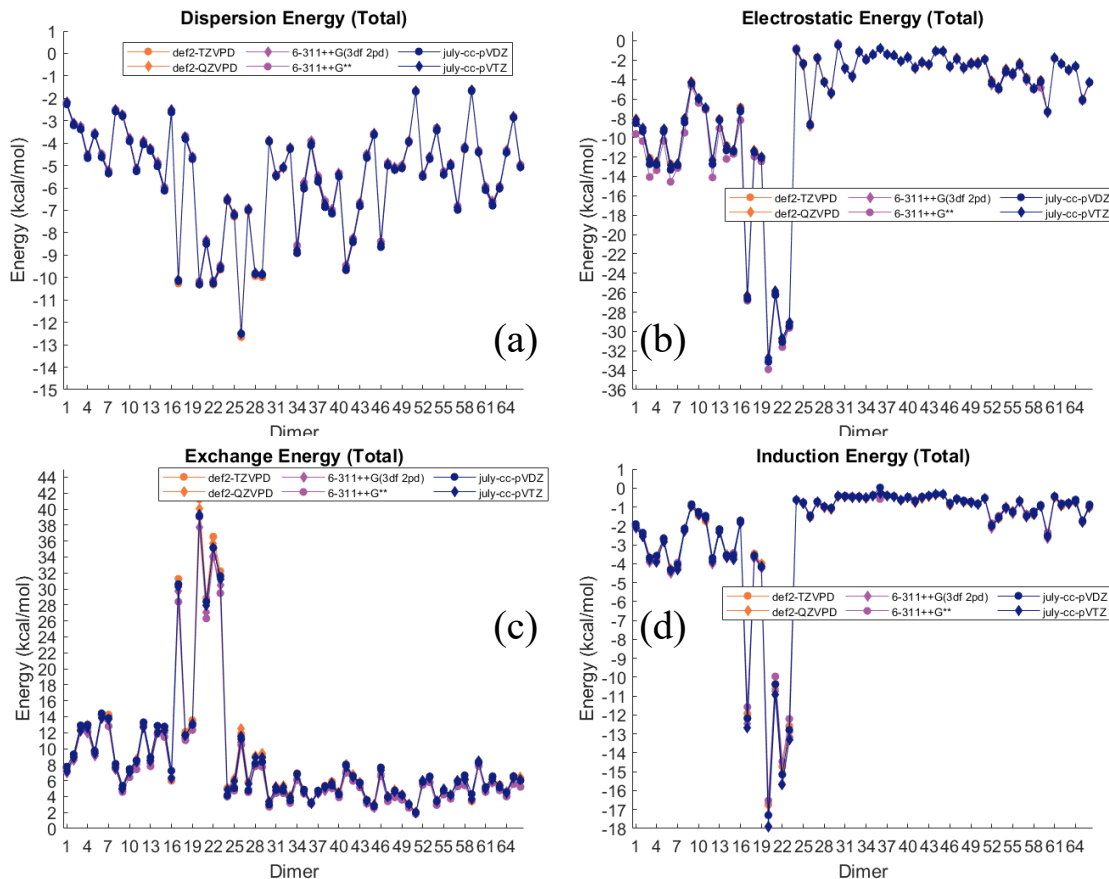


Fig. 3: XSAPT + MBD energy components evaluated using various basis sets. The energy components are (a) dispersion energy computed from the MBD model, which depends on the SCF monomer charge densities; (b) electrostatic energy, equal to the Coulomb interaction between isolated-monomer SCF charge densities; (c) exchange energy (*i.e.*, Pauli repulsion); and (d) induction energy.

best-performing basis set is jun-cc-pVTZ with a MAE of 0.8 kcal/mol although the ha-def2-SVPD and def2-TZVPPD basis sets afford comparable errors. Unlike the case for the S66 data set, where all of the monomers are charge-neutral, we see no systematic improvement of these errors as larger basis sets are employed. For these challenging systems, the benchmark study in Ref. 76 concluded that not just a δ HF but actually a δ MP2 calculation may be required to obtain CCSD(T)-quality accuracy, so the XSAPT + *ai*D3 error statistics for larger basis sets such as def2-QZVPD may simply reflect the practical limitations of the accuracy of this approach. Some error cancelation is then responsible for the somewhat better performance of the smaller basis sets mentioned above.

Figure 4 shows the energy components (excluding dispersion) for the IL16 data set, using XSAPT + *ai*D3 calculations in various basis sets. The electrostatics term (Fig. 4b) is especially flat, with variations of no more than 1 kcal/mol across a wide range of Karlsruhe and Dunning basis sets. Exchange energies (Fig. 4c) also vary by only ~ 1 kcal/mol if double- ζ basis sets are excluded from the

comparison. Induction energies span a wider range, up to 4 kcal/mol if double- ζ basis sets are again excluded. This points to induction as the energy component wherein the overall errors in the interaction energies reside, which is consistent with the need for a δ MP2 correction.

In contrast to IL16, errors in the interaction energies for the AHB21 and CHB6 data sets follow more discernible and systematic trends; see Table 3. The smallest MAEs are obtained using the aug-cc-pVQZ and def2-QZVPD basis sets, the latter of which is essentially identical in its performance to def2-QZVPPD. Cationic systems in CHB6 afford slightly lower MAEs when compared to the anionic systems in AHB21 but the basis-set trends are similar. In these systems where only a single partner species is charged, the trend is that the more diffuse basis sets afford the smallest errors and pruning the diffuse functions has a generally detrimental effect.

E. Hybrid Calculation of the Interaction Energy

Finally we consider whether a hybrid method can be used in which different energy components are computed at

Table 3: Error Statistics (in kcal/mol) for XSAPT + *ai*D3 Applied to Ion-Containing Data Sets.^a

Method	IL16			AHB21			CHB6		
	MAE ^b	MPE ^c	Max ^d	MAE ^b	MPE ^c	Max ^d	MAE ^b	MPE ^c	Max ^d
aug-cc-pVDZ	1.29	1.20	3.54	1.86	8.05	7.36	2.01	6.19	6.50
jul-cc-pVDZ	1.09	1.03	3.18	2.26	11.00	7.69	2.05	6.31	6.19
jun-cc-pVDZ	2.53	2.31	5.75	3.12	12.88	13.12	1.46	5.04	3.11
aug-cc-pVTZ	1.71	1.57	3.83	1.27	5.76	5.70	2.09	6.32	7.06
jul-cc-pVTZ	1.28	1.18	3.46	1.48	7.03	5.69	2.32	7.11	7.57
jun-cc-pVTZ	0.80	0.75	3.07	1.76	8.45	6.44	2.09	6.39	7.21
aug-cc-pVQZ	1.72	1.57	3.69	1.23	5.52	8.18	1.31	3.83	4.74
jul-cc-pVQZ	1.44	1.31	3.38	1.28	6.17	7.80	1.69	5.14	5.53
jun-cc-pVQZ	1.17	1.08	2.79	1.43	6.91	6.18	1.53	4.78	4.40
def2-SVPD	1.23	1.14	2.60	2.49	11.31	8.14	1.52	5.31	2.89
ha-def2-SVPD	0.88	0.83	3.38	2.45	11.64	9.05	1.38	4.92	2.59
ma-def2-SVPD	3.87	3.54	7.90	3.35	16.45	10.31	1.63	6.57	3.65
def2-TZVPD	1.08	0.99	2.29	2.54	11.09	8.36	1.26	4.01	3.65
def2-TZVPPD	0.85	0.80	2.16	1.89	8.87	6.09	1.63	5.00	5.78
ha-def2-TZVPD	1.37	1.26	2.31	2.82	12.38	8.43	1.33	4.25	3.90
ma-def2-TZVPD	2.24	2.05	4.37	2.41	10.61	7.69	1.93	6.74	3.99
def2-QZVPD	1.24	1.14	2.63	1.46	7.28	3.06	1.20	3.75	3.96
def2-QZVPPD	1.24	1.14	2.63	1.46	7.28	3.06	1.21	3.78	3.96
ha-def2-QZVPD	1.23	1.14	2.56	1.55	7.99	3.33	1.43	4.44	4.87
ma-def2-QZVPD	1.43	1.31	3.37	1.29	6.53	3.61	1.89	5.75	6.13

^aBoldface values indicate the best-performing Dunning and Karlsruhe basis sets^bMean absolute error^cMean percent error^dMaximum error

different levels of theory, exploiting the separability of the SAPT decomposition. A simple-to-implement version of such a procedure is to focus on the δE_{HF} correction, which is the only part of XSAPT + MBD or XSAPT + *ai*D that requires an iterative SCF procedure to be performed on a dimer. This is usually the computational bottleneck step when high-quality basis sets are used. Furthermore, the energy difference that defines δE_{HF} in eq. 4 suggests that this term might converge to the CBS limit more rapidly than other energy components in eq. 1.

To test this, we repeated the XSAPT + MBD/def2-QZVPD calculations on the S66 dimers but evaluated δE_{HF} using a smaller basis set for both the supramolecular HF calculation and the coupled-perturbed HF equations that are needed to evaluate the induction terms in eq. 4. (The larger basis set is used for all of the terms in eq. 1 except δE_{HF} .) Figure 5 shows the absolute errors in S66 interaction energies for this hybrid calculation, using various choices for the smaller basis set. The use of basis sets as small as ma-def2-SVPD affects the accuracy by $\lesssim 1$ kcal/mol.

As shown by the timing data in Fig. 6, these hybrid approaches have a dramatic effect on the calculation time required. Timings in Fig. 6 represent average values across the S66 set of dimers, where all terms in the XSAPT + MBD interaction except for δE_{HF} are evaluated using the def2-QZVPD basis set. If the δE_{HF}

correction is evaluated instead using ma-def2-SVPD, the resulting composite calculation is $81\times$ faster (wall time on a single compute node) compared to a conventional calculation that uses the quadruple- ζ basis set for the δE_{HF} correction as well. The reduction in total compute time (across all cores) is $\approx 150\times$ for this combination of basis sets, which according to Fig. 5 incurs < 1 kcal/mol error.

5 Conclusions

It has been shown that the different energy components converge at different rates when it comes to system size. A general trend across all of the different components is that diffuse orbitals are important, and if they are neglected, the accuracy in the calculation is drastically decreased. The electrostatic energy is sensitive to diffuse functions used in the basis set, and if they are used, it is shown to converge very early and gives high accuracy even at the def2-SVPD level. The exchange energy is one of the most sensitive energy component and does not converge until the triple- ζ level. The induction, like the exchange energy is also very sensitive to basis set used. It is shown to not converge until the triple- ζ level as well. Finally, the dispersion, at least in the XSAPT+MBD formalism, converges quickly and reaches near convergence

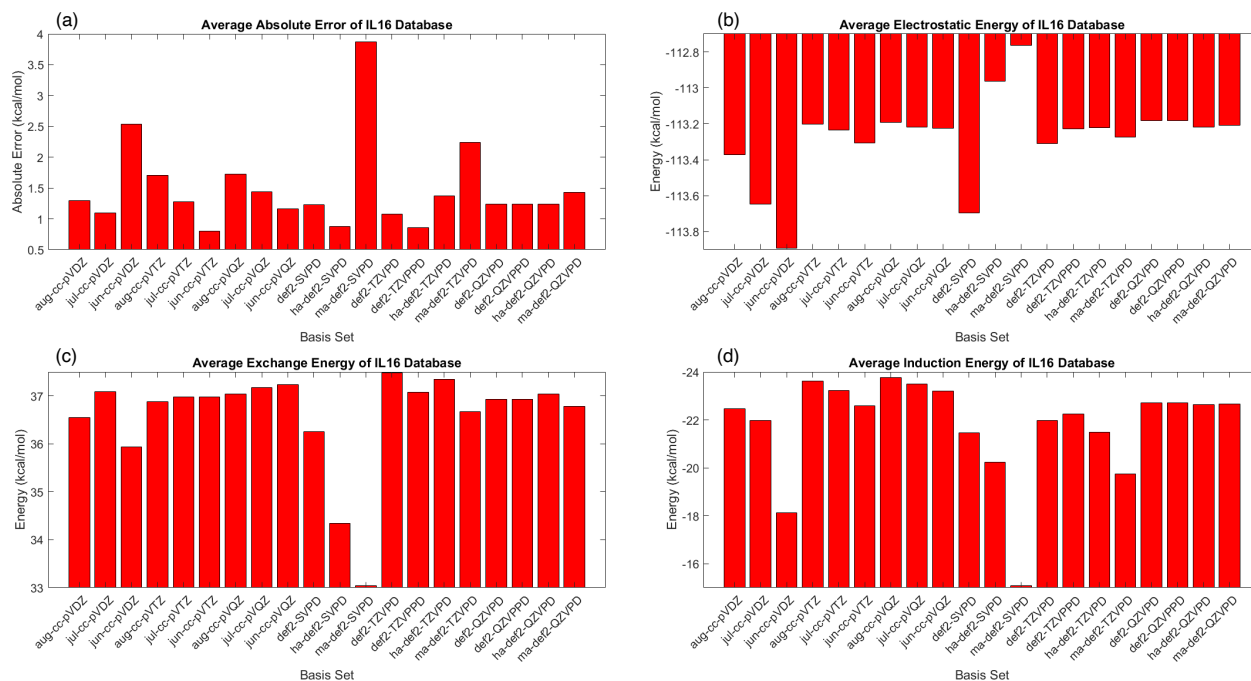


Fig. 4: Energy components (excluding dispersion) for the IL16 data set, evaluated using XSAPT + *ai*D3 in various basis sets.

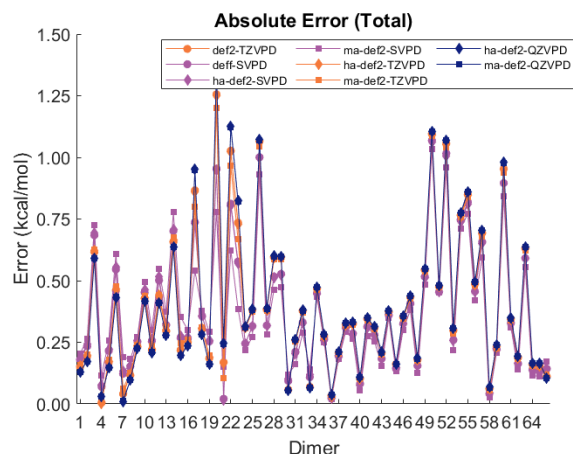


Fig. 5: Absolute errors across the S66 data set for XSAPT + MBD/def2-QZVPPD calculations in which the δ HF term is computed using a smaller basis set.

at the double- ζ level.

It has also been shown that the energy decomposition

analysis properties of SAPT allows for the different energy components to be calculated at different levels of theory. In the case of this paper, the δ HF calculation was shown that it can be done in much smaller basis sets than the other portions of the SAPT calculation without causing any significant changes in the accuracy of the calculation. The computational speed ups from this type of calculation are on the level of 80 times (wall time) to 150 times (CPU time). This aids in reducing the computational times of the calculations and allows for larger systems to be studied while also keeping high accuracy in the method.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550. Calculations were performed at the Ohio Supercomputer Center under project no. PAA-0003.⁷⁹

¹ Szalewicz, K. Determination of structure and properties of molecular crystals from first principles. *Acc. Chem. Res.* **2014**, *47*, 3266–3274.

² Brandenburg, J. G.; Gerit, J.; Grimme, S. Dispersion corrected Hartree-Fock and density functional theory for organic crystal structure prediction. In *Prediction and Calcula-*

tion of Crystal Structures, Vol. 345; Ş. Atahan-Evrenk; Aspuru-Guzik, A., Eds.; Springer: Heidelberg, 2014.

³ Beran, G. J. O. Modeling polymorphic molecular crystals with electronic structure theory. *Chem. Rev.* **2016**, *116*, 5567–5613.

⁴ Beran, G. J. O.; Heit, Y. N.; Hartman, J. D. Nonco-

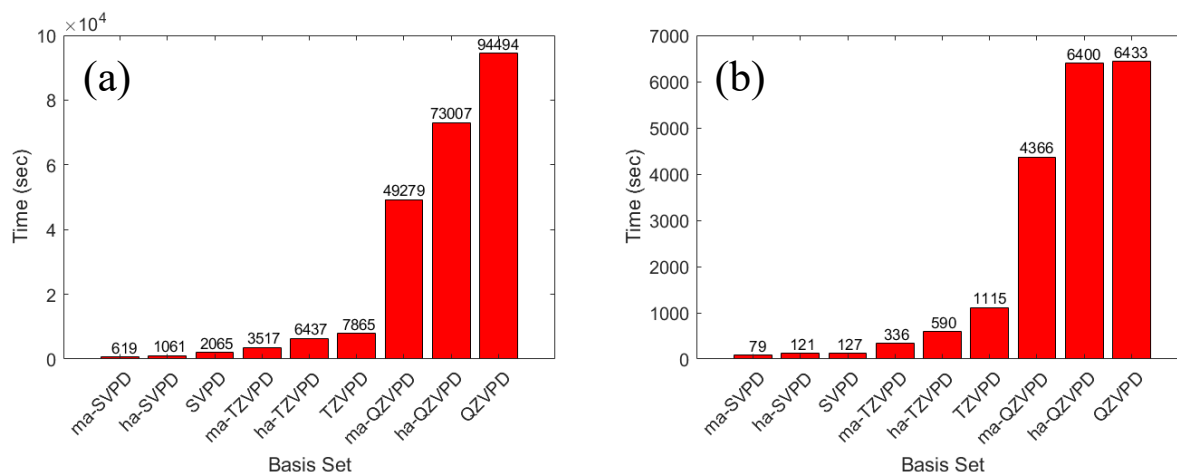


Fig. 6: (a) Total CPU time versus (b) wall time required to evaluate interaction energies at the XSAPT + MBD/def2-QZVPD level of theory using a hybrid approach in which the δ HF correction is evaluated in one of the smaller basis sets that is listed along the horizontal axis. Timings are averages over all of the S66 dimers and all calculations were performed on a single 28-core node.

valent interactions in molecular crystals. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 10, pages 303–331.

- ⁵ Price, S. L.; Brandenburg, J. G. Molecular crystal structure prediction. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 11, pages 333–363.
- ⁶ Plevin, M.; Bryce, D. L.; Boisbouvier, J. Direct detection of CH/ π interactions in proteins. *Nat. Chem.* **2010**, *2*, 466–471.
- ⁷ McGaughey, G. B.; Gagné, M.; Rappé, A. K. π -stacking interactions: Alive and well in proteins. *J. Biol. Chem.* **1998**, *273*, 15458–15463.
- ⁸ Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. C-H $\cdots\pi$ interactions in proteins. *J. Mol. Biol.* **2001**, *307*, 357–377.
- ⁹ He, X.; Fusti-Molnar, L.; Cui, G.; Merz, Jr., K. M. The importance of dispersion and electron correlation in ab initio protein folding. *J. Phys. Chem. B* **2009**, *113*, 5290–5300.
- ¹⁰ Xu, Z.; Zhang, Q.; Shi, J.; Zhu, W. Underestimated noncovalent interactions in protein data bank. *J. Chem. Inf. Model* **2019**, *59*, 3389–3399.
- ¹¹ Parrish, R. M.; Sitkoff, D. F.; Cheney, D. L.; Sherrill, C. D. The surprising importance of peptide bond contacts in drug–protein interactions. *Chem. Eur. J.* **2017**, *23*, 7887–7890.
- ¹² Kumar, K.; Woo, S. M.; Siu, T.; Cortopassi, W. A.; Duarte, F.; Paton, R. S. Cation– π interactions in protein–ligand binding: Theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* **2018**, *9*, 2655–2665.
- ¹³ Sylvetsky, N. Toward simple, predictive understanding of protein–ligand interactions: Electronic structure calculations on Torpedo Californica Acetylcholinesterase join forces with the chemist’s intuition. *Sci. Rep.* **2020**, *10*, 9219:1–12.
- ¹⁴ Schirber, J. B.; Nascimento, D. R.; Koutsoukas, A.; Spronk, S. A.; Cheney, D. L.; Sherrill, C. D. CLIFF: A component-based, machine-learned, intermolecular force field. *J. Chem. Phys.* **2021**, *154*, 184110:1–17.
- ¹⁵ Tkatchenko, A. Current understanding of van der Waals effects in realistic materials. *Adv. Funct. Mater.* **2015**, *25*, 2054–2061.
- ¹⁶ Ravva, M. K.; Risko, C.; Brédas, J.-L. Noncovalent interactions in organic electronic materials. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 9, pages 277–302.
- ¹⁷ Cooper, V. R.; Lam, C. N.; Wang, Y.; Sumpter, B. G. Noncovalent interactions in nanotechnology. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 14, pages 417–451.
- ¹⁸ Wagner, J. P.; Schreiner, P. R. London dispersion in molecular chemistry—Reconsidering steric effects. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 12274–1229.
- ¹⁹ Herbert, J. M. Neat, simple, and wrong: Debunking electrostatic fallacies regarding noncovalent interactions. *J. Phys. Chem. A* **2021**, *125*, 7125–7137.
- ²⁰ Carter-Fenk, K.; Lao, K. U.; Herbert, J. M. Predicting and understanding non-covalent interactions using novel forms of symmetry-adapted perturbation theory. *Acc. Chem. Res.* **2021**, *54*, 3679–3690.
- ²¹ Szalewicz, K.; Patkowski, K.; Jezierski, B. Intermolecular interactions via perturbation theory: From diatoms to biomolecules. In *Intermolecular Forces and Clusters II*, Vol. 116; Wales, D. J., Ed.; Springer-Verlag: Berlin, 2005.
- ²² Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *WIREs Comput. Mol. Sci.* **2012**, *2*, 304–326.
- ²³ Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *WIREs Comput. Mol. Sci.* **2012**, *2*, 254–272.
- ²⁴ Jansen, G. Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions.

- WIREs Comput. Mol. Sci.* **2014**, *4*, 127–144.
- 25 Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140*, 094106:1–16.
 - 26 Patkowski, K. Recent developments in symmetry-adapted perturbation theory. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1452:1–47.
 - 27 Stone, A. J. Physical basis of intermolecular interactions. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 1, pages 3–26.
 - 28 Francisco, E.; Pendás, A. M. Energy partition analyses: Symmetry-adapted perturbation theory and other techniques. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 2, pages 27–64.
 - 29 Misquitta, A. Intermolecular interactions. In *Handbook of Computational Chemistry*, 2nd ed.; Leszczynski, J.; Kaczmarek-Kedziera, A.; Puzyn, T.; Papadopoulos, M. G.; Reis, H.; Shukla, M. K., Eds.; Springer International Publishing: Switzerland, 2017; Chapter 8, pages 295–335.
 - 30 Hohenstein, E. G.; Duan, J.; Sherrill, D. C. Origin of the surprising enhancement of electrostatic energies by electron-donating substituents in substituted sandwich benzene dimers. *J. Am. Chem. Soc.* **2011**, *133*, 13244–13247.
 - 31 Sherrill, C. D. Energy component analysis of π interactions. *Acc. Chem. Res.* **2013**, *46*, 1020–1028.
 - 32 Parrish, R. M.; Sherrill, C. D. Quantum-mechanical evaluation of the π – π versus substituent– π interactions in π stacking: Direct evidence for the Wheeler–Houk picture. *J. Am. Chem. Soc.* **2014**, *136*, 17386–17389.
 - 33 Parrish, R. M.; Parker, T. M.; Sherrill, C. D. Chemical assignment of symmetry-adapted perturbation theory interaction energy components: The functional-group SAPT partition. *J. Chem. Theory Comput.* **2014**, *10*, 4417–4431.
 - 34 Carter-Fenk, K.; Herbert, J. M. Electrostatics does not dictate the slip-stacked arrangement of aromatic π – π interactions. *Chem. Sci.* **2020**, *11*, 6758–6765.
 - 35 Carter-Fenk, K.; Herbert, J. M. Reinterpreting π -stacking. *Phys. Chem. Chem. Phys.* **2020**, *22*, 24870–24886.
 - 36 Herbert, J. M.; Carter-Fenk, K. Electrostatics, charge transfer, and the nature of the halide–water hydrogen bond. *J. Phys. Chem. A* **2021**, *125*, 1243–1256.
 - 37 Shahbaz, M.; Szalewicz, K. Do semilocal density-functional approximations recover dispersion energies at small intermonomer separations? *Phys. Rev. Lett.* **2018**, *121*, 113402:1–6.
 - 38 Shahbaz, M.; Szalewicz, K. Evaluation of methods for obtaining dispersion energies used in density functional calculations of intermolecular interactions. *Theor. Chem. Acc.* **2019**, *138*, 25:1–17.
 - 39 Price, A. J. A.; Bryenton, K. R.; Johnson, E. R. Requirements for an accurate dispersion-corrected density functional. *J. Chem. Phys.* **2021**, *154*, 230902:1–12.
 - 40 Schmidt, J. R.; Yu, K.; McDaniel, J. G. Transferable next-generation force fields from simple liquids to complex materials. *Acc. Chem. Res.* **2015**, *48*, 548–556.
 - 41 Metz, M. P.; Szalewicz, K. Automatic generation of flexible-monomer intermolecular potential energy surfaces. *J. Chem. Theory Comput.* **2020**, *16*, 2317–2339.
 - 42 Gray, M.; Herbert, J. M. Simplified tuning of long-range corrected density functionals for symmetry-adapted perturbation theory. *J. Chem. Phys.* **2021**, *155*, 034103:1–8.
 - 43 Lao, K. U.; Herbert, J. M. Symmetry-adapted perturbation theory with Kohn–Sham orbitals using non-empirically tuned, long-range-corrected density functionals. *J. Chem. Phys.* **2014**, *140*, 044108:1–8.
 - 44 Hapka, M.; Rajchel, L.; Modrzejewski, M.; Chalański, G.; Szczęśniak, M. M. Tuned range-separated hybrid functionals in the symmetry-adapted perturbation theory. *J. Chem. Phys.* **2014**, *141*, 134120:1–10.
 - 45 Szabo, A.; Ostlund, N. S. The correlation energy in the random phase approximation: Intermolecular forces between closed-shell systems. *J. Chem. Phys.* **1977**, *67*, 4351–4360.
 - 46 Heßelmann, A. Improved supermolecular second order Møller–Plesset intermolecular interaction energies using time-dependent density functional response theory. *J. Chem. Phys.* **2008**, *128*, 144112:1–9.
 - 47 Řezáč, J.; Greenwell, C.; Beran, G. J. O. Accurate noncovalent interactions via dispersion-corrected second-order Møller–Plesset perturbation theory. *J. Chem. Theory Comput.* **2018**, *14*, 4711–4721.
 - 48 Lao, K. U.; Herbert, J. M. Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes. *J. Chem. Theory Comput.* **2018**, *14*, 2955–2978.
 - 49 Nguyen, B.; Chen, G. P.; Agee, M. M.; Burow, A. M.; Tang, M.; Furche, F. Divergence of many-body perturbation theory for noncovalent interactions of large molecules. *J. Chem. Theory Comput.* **2020**, *16*, 2258–2273.
 - 50 Heßelmann, A.; Jansen, G.; Schütz, M. Density-functional theory symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005**, *122*, 014103:1–17.
 - 51 Bukowski, R.; Podeszwa, R.; Szalewicz, K. Efficient calculation of coupled Kohn–Sham dynamic susceptibility functions and dispersion energies with density fitting. *Chem. Phys. Lett.* **2005**, *414*, 111–116.
 - 52 Lao, K. U.; Herbert, J. M. Accurate intermolecular interactions at dramatically reduced cost: XPol+SAPT with empirical dispersion. *J. Phys. Chem. Lett.* **2012**, *3*, 3241–3248.
 - 53 Lao, K. U.; Herbert, J. M. An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method. *J. Chem. Phys.* **2013**, *139*, 034107:1–16. Erratum: *J. Chem. Phys.* **140**, 119901 (2014).
 - 54 Lao, K. U.; Herbert, J. M. Accurate and efficient quantum chemistry calculations of noncovalent interactions in many-body systems: The XSAPT family of methods. *J. Phys. Chem. A* **2015**, *119*, 235–253.
 - 55 Carter-Fenk, K.; Lao, K. U.; Liu, K.-Y.; Herbert, J. M. Accurate and efficient *ab initio* calculations for supramolecular complexes: Symmetry-adapted perturbation theory with many-body dispersion. *J. Phys. Chem. Lett.* **2019**, *10*, 2706–2714.
 - 56 Liu, K.-Y.; Carter-Fenk, K.; Herbert, J. M. Self-consistent charge embedding at very low cost, with application to symmetry-adapted perturbation theory. *J. Chem. Phys.* **2019**, *151*, 031102:1–7.
 - 57 Jacobson, L. D.; Herbert, J. M. Polarization-bound quasi-

- continuum states are responsible for the ‘blue tail’ in the optical absorption spectrum of the aqueous electron. *J. Am. Chem. Soc.* **2010**, *132*, 10000–10002.
- ⁵⁸ Herbert, J. M.; Jacobson, L. D.; Lao, K. U.; Rohrdanz, M. A. Rapid computation of intermolecular interactions in molecular and ionic clusters: Self-consistent polarization plus symmetry-adapted perturbation theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7679–7699.
- ⁵⁹ Jacobson, L. D.; Richard, R. M.; Lao, K. U.; Herbert, J. M. Efficient monomer-based quantum chemistry methods for molecular and ionic clusters. *Annu. Rep. Comput. Chem.* **2013**, *9*, 25–58.
- ⁶⁰ Herbert, J. M.; Paul, S. K. Interaction energy analysis of monovalent inorganic anions in bulk water versus air/water interface. (DOI: 10.33774/chemrxiv-2021-xrht0-v2)
- ⁶¹ Moszyński, R.; Cybulski, S. M.; Chalaśiński, G. Many-body theory of intermolecular induction interactions. *J. Chem. Phys.* **1994**, *100*, 4998–5010.
- ⁶² Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states. *J. Chem. Phys.* **2009**, *130*, 054112:1–8.
- ⁶³ Modrzejewski, M.; Rajchel, L.; Chalaśiński, G.; Szczesniak, M. M. Density-dependent onset of the long-range exchange: A key to donor–acceptor properties. *J. Phys. Chem. A* **2013**, *117*, 11580–11586.
- ⁶⁴ Ambrosetti, A.; Reilly, A. M.; DiStasio Jr., R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **2014**, *140*, 18A508:1–14.
- ⁶⁵ Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.
- ⁶⁶ Carter-Fenk, K.; Mundy, C. J.; Herbert, J. M. Natural charge-transfer analysis: Eliminating spurious charge-transfer states in time-dependent density functional theory via diabaticization, with application to projection-based embedding. *J. Chem. Theory Comput.* **2021**, *17*, 4195–4210.
- ⁶⁷ Jensen, F. Atomic orbital basis sets. *WIREs Comput. Mol. Sci.* **2012**, *3*, 273–295.
- ⁶⁸ Epifanovsky, E. *et al.* Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801:1–59.
- ⁶⁹ Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- ⁷⁰ Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105:1–11.
- ⁷¹ Dunning, Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- ⁷² Kendall, R. A.; Dunning, Jr., T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- ⁷³ Zheng, E. P. J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034.
- ⁷⁴ Zheng, J.; Xu, X.; Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **2011**, *128*, 295–305.
- ⁷⁵ Řezáč, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438. Erratum: *J. Chem. Theory Comput.* **10**, 1359–1360 (2014).
- ⁷⁶ Lao, K. U.; Schäffer, R.; Jansen, G.; Herbert, J. M. Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory. *J. Chem. Theory Comput.* **2015**, *11*, 2473–2486.
- ⁷⁷ Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364–3374.
- ⁷⁸ Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269.
- ⁷⁹ Ohio Supercomputer Center, <http://osc.edu/ark:/19495/f5s1ph73>.